

Theory of Parallel Particle Filters for Hidden Markov Models

BY HOCK PENG CHAN, CHIANG WEE HENG & AJAY JASRA

Department of Statistics & Applied Probability, National University of Singapore, Singapore, 117546, SG.

E-Mail: *stachp@nus.edu.sg*, *a0068363@nus.edu.sg*, *staja@nus.edu.sg*

Abstract

The objective of this article is to study the asymptotic behavior of a new particle filtering approach in the context of hidden Markov models (HMMs). In particular, we develop an algorithm where the latent-state sequence is segmented into multiple shorter portions, with an estimation technique based upon a separate particle filter in each portion. The partitioning facilitates the use of parallel processing. Based upon this approach, we introduce new estimators of the latent states and likelihood which have similar or better variance properties compared to estimators derived from standard particle filters. Moreover due to parallelization there is less wall-clock computational time. We show that the likelihood function estimator is unbiased, prove central limit theorem convergences of estimators, and provide consistent in-sample estimation of the asymptotic variances. The theoretical analyses, supported by a numerical study, show that the segmentation reduces the variances in smoothed latent-state estimators, in addition to the savings in wall-clock time.

Key Words: CLT; parallel processing; SMC; standard error estimation.

1 Introduction

Hidden Markov models are a flexible class of statistical models that are applied in a wide variety of applications such as bioinformatics, economics, engineering and finance; see [5] for an introduction. Mathematically a HMM corresponds to a pair of discrete-time processes $X_t \in \mathbf{X}$, $Y_t \in \mathbf{Y}$, with the observed Y_t conditionally independent given X_t and the unobserved X_t obeying a first-order Markov chain

$$(X_t|X_{t-1} = x) \sim P_\theta(\cdot|x), \quad (Y_t|X_t = x) \sim G_\theta(\cdot|x), \quad t \geq 1 \quad (1.1)$$

with densities, p_θ and g_θ with respect to dominating measures on their state-spaces and θ is a static parameter.

From an inferential perspective, we are interested in the conditional distribution of X_t given all the observations up to and perhaps after time t . This has a wide-range of interpretations, particularly in real-time applications. In addition there is much practical interest in the calculation of the likelihood of the observations, for model comparison and parameter estimation. The difficulty with the afore-mentioned objectives is that the exact computation of the conditional distribution or likelihood is typically not possible, as often the high-dimensional integrals that it depends on are often difficult to evaluate. In practice Monte Carlo-based numerical methods are adopted, in particular the method of particle filters or equivalently sequential Monte Carlo (SMC); see [12] for an overview.

SMC methods can be described as a collection of techniques that approximate a sequence of distributions, known up to normalizing constants, of increasing dimensions, and are often applied to HMMs. SMC methods combine importance sampling and resampling to approximate distributions. The idea is to introduce a sequence of proposal densities and sequentially simulate a collection of $K > 1$ samples, termed particles, in parallel from these proposals. In most scenarios it is not possible to use the distribution of interest as a proposal. Therefore one must correct for the discrepancy between proposal and target via importance weights. As the variance of these importance weights can potentially increase exponentially with algorithmic time, resampling is applied to control it. Resampling consists of sampling with replacement from the current samples using the weights and then resetting them to K^{-1} . The theoretical properties of SMC with regards to their convergence as K grows are well-studied; see [6, 8, 10, 11, 15].

In recent years, the applicability of SMC techniques has been enhanced by parallel computation; see [17]. One of the main bottlenecks in the application of parallel computation to SMC methods is the resampling step, a major requirement for the method to be efficient. This has led to a number of researchers investigating methodologies that reduce the degree of interaction in SMC algorithms; see [16, 18, 21]. This work is complementary to the afore-mentioned references, and is a methodology designed to assist in the parallelization of SMC algorithms, while attempting to retain their attractive properties. Our objective is to study

the asymptotic behavior of HMM estimators when the latent-state sequence is segmented into multiple shorter portions, by applying an estimation technique based upon a separate particle filter in each portion. The partitioning facilitates the use of parallel processing. Based upon this approach, we introduce new SMC-based estimators of the latent states (that is, expectations w.r.t. the filter and smoother) and likelihood with similar or better variance properties compared to standard SMC estimators, but due to parallelization can be calculated in less wall-clock computational time. In particular we show:

- unbiasedness of our likelihood estimate,
- central limit convergences of the likelihood and latent-state estimates,
- consistent estimation of asymptotic variances.

Our likelihood estimates can be used in conjunction with recent advances in SMC methodology in which particle filter processing is just one component of a two-layered process when learning θ in a Bayesian manner: particle MCMC (PMCMC) [2], SMC² [9], MCMC substitution [7]. That is, our procedure can be routinely used in the context of these works. In principle, there is no need to break up the observation sequence into strictly disjoint segments, it can be advantageous to include additional observations at the edges of the segments to smooth out the joining of the sample paths. We shall illustrate this in the numerical study in Section 3.3.2.

We describe the algorithm and estimators in Section 2 and the asymptotic theory in Section 3, with an illustration of variance reduction in smoothed latent-state estimators. We discuss refinements in Section 4. The technical proofs are consolidated in the Appendix.

2 Independent particle filters for segmented data

Let $\mathbf{Y}_U = (Y_1, \dots, Y_U)$ for some $U > 1$. As the observation sequence is conditionally independent given the latent-state sequence, the density of $\mathbf{X}_U := (X_1, \dots, X_U)$ conditioned on \mathbf{Y}_U is given by

$$p_\theta(\mathbf{x}_U | \mathbf{Y}_U) = \prod_{t=1}^U [p_\theta(x_t | x_{t-1}) g_\theta(Y_t | x_t)] / \lambda(\theta), \quad (2.1)$$

where $\lambda(\theta) [= \lambda(\mathbf{Y}_U|\theta)]$ is the likelihood function that normalizes $p_\theta(\cdot|\mathbf{Y}_U)$ so that it integrates to 1, and $p_\theta(x_t|x_{t-1})$ for $t = 1$ denotes $p_\theta(x_1)$.

Let $\mathbf{x}_t = (x_1, \dots, x_t)$ and let $q_t(\cdot|\mathbf{x}_{t-1})$ be an importance density of X_t , with $q_t(\cdot|\mathbf{x}_{t-1})$ for $t = 1$ denoting $q_1(\cdot)$. We shall require that $q_t(x_t|\mathbf{x}_{t-1}) > 0$ whenever $p_\theta(x_t|x_{t-1}) > 0$. For notational simplicity, we assume that $U = MT$ for positive integers M and T , so that the latent-state sequence can be partitioned neatly into M subsequences of equal length T . We shall operate M particle filters independently, with the m th particle filtering generating sample paths of $\mathbf{X}_{m,mT}$, where $\mathbf{X}_{m,t} = (X_{(m-1)T+1}, \dots, X_t)$. Due to the independent nature of the particle filters, we require that for $(m-1)T < t \leq mT$, $q_t(\cdot|\mathbf{x}_{t-1})$ does not depend on $\mathbf{x}_{(m-1)T}$. We can thus express $q_t(\cdot|\mathbf{x}_{t-1})$ as $q_t(\cdot|\mathbf{x}_{m,t-1})$.

Let $w_t(\mathbf{x}_t)$ be the positive resampling weights of a sample path \mathbf{x}_t , and again due to the independent nature of the particle filters, we shall require that for $(m-1)T < t \leq mT$, $w_t(\mathbf{x}_t)$ does not depend on $\mathbf{x}_{(m-1)T}$, and express $w_t(\mathbf{x}_t)$ also as $w_t(\mathbf{x}_{m,t})$.

2.1 Approach

We shall apply standard multinomial resampling at every stage, as proposed in the seminal paper [14]; this is not necessary from a methodological point of view, but we will analyze this case. In the case of a single particle filter, it is common to adopt

$$q_t(x_t|\mathbf{x}_{t-1}) = p_\theta(x_t|x_{t-1}), \quad w_t(x_t) = g_\theta(Y_t|x_t),$$

but it need not be the case, and we can in general let

$$w_t(\mathbf{x}_t) = \frac{g_\theta(Y_t|x_t)p_\theta(x_t|x_{t-1})}{q_t(x_t|\mathbf{x}_{t-1})}. \quad (2.2)$$

Therefore the single particle filter targets (up-to proportionality), after resampling at stage t :

$$\prod_{u=1}^t [g_\theta(Y_u|x_u)p_\theta(x_u|x_{u-1})].$$

For the parallel particle filters, the m th particle filter, after resampling at stage $(m-1)T < t \leq mT$, targets

$$\begin{aligned} \pi_{m,t}(\mathbf{x}_{m,t}) &\propto r_m(x_{(m-1)T+1})g_\theta(Y_{(m-1)T+1}|x_{(m-1)T+1}) \\ &\times \prod_{u=(m-1)T+2}^t [g_\theta(Y_u|x_u)p_\theta(x_u|x_{u-1})], \end{aligned} \quad (2.3)$$

where $r_m(\cdot)$ is a positive probability on \mathbf{X} which can be evaluated up to a constant, and is independent from output of the others filters; sensible choices of $r_m(\cdot)$ are suggested in Section 3.3.2. For $m = 1$, we can simply let $r_m(x_1) = p_\theta(x_1)$. The forms of the target and proposal mean that the particle filters can be run in parallel with each other. Therefore w_t has the form given in (2.2), with the exceptions that when $t = (m-1)T + 1$, $p_\theta(x_t|x_{t-1})$ is replaced by $r_m(x_t)$.

The particle filter approach is given below. It is remarked that some of the notations, for example H_t^k , are not needed to run the particle filter but will help to facilitate the subsequent theoretical analysis. For $1 \leq m \leq M$:

Particle filter m (PFm). Recursively at stages $t = (m-1)T + 1, \dots, mT$:

1. *Importance sampling.* For $1 \leq k \leq K$, sample $\tilde{X}_t^k \sim q_t(\cdot|\mathbf{X}_{m,t-1}^k)$ and let $\tilde{\mathbf{X}}_{m,t}^k = \tilde{X}_t^k$ if $t = (m-1)T + 1$, $\tilde{\mathbf{X}}_{m,t}^k = (\mathbf{X}_{m,t-1}^k, \tilde{X}_t^k)$ otherwise.
2. *Resampling.* Generate i.i.d. B_t^1, \dots, B_t^K [$B(1), \dots, B(K)$ for short] such that

$$P\{B(1) = j\} = w_t(\tilde{\mathbf{X}}_{m,t}^j)/(K\bar{w}_t)(= W_t^j), \quad (2.4)$$

where $\bar{w}_t = K^{-1} \sum_{k=1}^K w_t(\tilde{\mathbf{X}}_{m,t}^k)$.

3. *Updating.* Let $(\mathbf{X}_{m,t}^k, A_{m,t}^k) = (\tilde{\mathbf{X}}_{m,t}^{B(k)}, A_{m,t-1}^{B(k)})$,

$$\tilde{H}_{m,t}^j = H_{m,t-1}^j/(KW_t^j) \text{ and } H_{m,t}^k = \tilde{H}_{m,t}^{B(k)}, \quad 1 \leq k \leq K, \quad (2.5)$$

with the conventions $A_{m,(m-1)T}^k = k$, $H_{m,(m-1)T}^k = 1$.

Remark 1. *There are other procedures that use two or more particle filters to perform estimation. For instance, [19] introduces a method based upon generalized two-filter smoothing. However, that approach is restricted to two particle filters that run forwards and backwards,*

and requires the choice of pseudo-densities, which may be more difficult than the choice of $r_m(\cdot)$. The approach of [20] uses multiple particle filters to perform estimation, but is different from the ideas in this article. Typically that approach will run filters in parallel on the same target and allow the filters themselves to interact. In our approach, we are able to reduce variability (relative to one particle filter) of estimates by segmentation, which is possibly not achieved in [20].

2.2 Notations

Set $\eta_0 = 1$ and assume that

$$\eta_t := E_q \left[\prod_{u=1}^t w_u(\mathbf{X}_u) \right] < \infty \text{ for } 1 \leq t \leq U, \quad (2.6)$$

where for (integrable) $\varphi : \mathbf{X}^t \rightarrow \mathbb{R}$,

$$E_q \varphi(\mathbf{X}_t) = \int_{\mathbf{X}^t} \varphi(\mathbf{x}_t) \left[\prod_{u=1}^t q_u(x_u | \mathbf{x}_{u-1}) \right] d\mathbf{x}_t. \quad (2.7)$$

Consider $(m-1)T < t \leq mT$. Define $\eta_{m,t} = \eta_t / \eta_{(m-1)T}$, and let

$$h_t(\mathbf{x}_t) = \eta_t \Big/ \prod_{u=1}^t w_u(\mathbf{x}_u), \quad h_{m,t}(\mathbf{x}_{m,t}) = \eta_{m,t} \Big/ \prod_{u=(m-1)T+1}^t w_u(\mathbf{x}_{m,u}). \quad (2.8)$$

By (2.4)–(2.8),

$$H_{m,t}^k = \left(\frac{\bar{w}_{(m-1)T+1} \cdots \bar{w}_t}{\eta_{m,t}} \right) h_{m,t}(\mathbf{X}_{m,t}^k). \quad (2.9)$$

Let $Z^m (= Z_K^m) = \{(k(1), \dots, k(m)) : 1 \leq k(n) \leq K \text{ for } 1 \leq n \leq m\}$. For $\mathbf{k} \in Z^m$, let

$$\begin{aligned} \tilde{\mathbf{X}}_t^{\mathbf{k}} &= (\mathbf{X}_{1,T}^{k(1)}, \dots, \mathbf{X}_{m-1,(m-1)T}^{k(m-1)}, \tilde{\mathbf{X}}_{m,t}^{k(m)}), & \tilde{H}_t^{\mathbf{k}} &= \left(\prod_{n=1}^{m-1} H_{n,nT}^{k(n)} \right) \tilde{H}_{m,t}^{k(m)}, \\ \mathbf{X}_t^{\mathbf{k}} &= (\mathbf{X}_{1,T}^{k(1)}, \dots, \mathbf{X}_{m-1,(m-1)T}^{k(m-1)}, \mathbf{X}_{m,t}^{k(m)}), & H_t^{\mathbf{k}} &= \left(\prod_{n=1}^{m-1} H_{n,nT}^{k(n)} \right) H_{m,t}^{k(m)}. \end{aligned} \quad (2.10)$$

Thus analogous to (2.9),

$$H_t^{\mathbf{k}} = \left(\frac{\bar{w}_1 \cdots \bar{w}_t}{\eta_t} \right) h_t(\mathbf{X}_t^{\mathbf{k}}). \quad (2.11)$$

The notation $A_{m,t}^k$ refers to the first-generation ancestor of $\mathbf{X}_{m,t}^k$ (or $\tilde{\mathbf{X}}_{m,t+1}^k$). That is $A_t^k = j$ if the first component of $\mathbf{X}_{m,t}^k$ is $\tilde{X}_{m-1,T+1}^j$. This ancestor tracing is exploited in Sections 3.3.3 for standard error approximations of the estimates. Finally $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2 .

3 Estimation theory

We are interested in the estimation of the likelihood $\lambda(\theta)$, and also of $\psi_U := E_p[\psi(\mathbf{X}_U)|\mathbf{Y}_U]$ for some real-valued measurable function ψ . Here E_p denotes expectation under the HMM (1.1). The estimation of $\lambda(\theta)$ falls under the canonical case; the theory is given in Section 3.2. The estimation of ψ_U falls under the non-canonical case; the theory is given in Section 3.3.

3.1 Estimates and Remarks

3.1.1 Canonical Case

Define the function

$$L(\mathbf{x}_U) = p_\theta(\mathbf{x}_U|\mathbf{Y}_U) / \prod_{t=1}^U q_t(x_t|\mathbf{x}_{t-1}), \quad (3.1)$$

where $p_\theta(\mathbf{x}_U|\mathbf{Y}_U)$ is as (2.1). The estimator of ψ_U in the canonical case, which we will prove is unbiased, is given by

$$\hat{\psi}_U = K^{-M} \sum_{\mathbf{k} \in \mathbb{Z}^M} L(\mathbf{X}_U^{\mathbf{k}}) \psi(\mathbf{X}_U^{\mathbf{k}}) H_U^{\mathbf{k}}. \quad (3.2)$$

By (2.1) and (3.1), $\lambda(\theta)$ appears in the denominator on the R.H.S. of (3.2). This does not pose a problem in the estimation of $\lambda(\theta)$, for by setting $\psi \equiv \lambda(\theta)$, we cancel out $\lambda(\theta)$. We define $\hat{\lambda}(\theta)$ to be the estimator obtained this way, that is

$$\hat{\lambda}(\theta)/\lambda(\theta) = K^{-M} \sum_{\mathbf{k} \in \mathbb{Z}^M} L(\mathbf{X}_U^{\mathbf{k}}) H_U^{\mathbf{k}}. \quad (3.3)$$

To further understand this estimate, we rewrite (3.3) as

$$\hat{\lambda}(\theta) = \left(\prod_{t=1}^U \bar{w}_t \right) \prod_{m=2}^M \left(K^{-2} \sum_{k=1}^K \sum_{\ell=1}^K \frac{p_\theta(X_{(m-1)T+1}^\ell | X_{(m-1)T}^k)}{r_m(X_{(m-1)T+1}^\ell)} \right), \quad (3.4)$$

where $X_{(m-1)T+1}^\ell$ here refers to the first component of $X_{m,mT}^\ell$. A heuristic justification is as follows. For expositional purposes, let us consider the simplest case of $M = 2$. The final term on the R.H.S. of (3.4), the double summation, is an SMC estimate of the ratio, up to a constant, of the actual target of interest (2.1), and the normalized target (2.3) that is sampled by the two particle filters. That is as $K \rightarrow \infty$, we would like to obtain convergence to

$$\int_{\mathbf{x}_U} \frac{p_\theta(x_{T+1}|x_T)}{r_2(x_{T+1})} \pi_{1,T}(\mathbf{x}_T) \pi_{2,U}(\mathbf{x}_{2,U}) d\mathbf{x}_U.$$

The term $(\prod_{t=1}^U \bar{w}_t)$ will converge in probability to the normalizing constants of $\pi_{2,U}(\mathbf{x}_{2,U})$.

The expression (3.4) also suggests a good choice of $r_2(\cdot)$. If we take

$$r_2(x) = K^{-1} \sum_{k=1}^K p_\theta(x|x_T^k),$$

then the double sum on the R.H.S. of (3.4) is exactly 1; that is, it does not contribute to the variance of the estimate. The choice above is exactly the SMC approximation of the predictor. However, the choice suggested above is not reasonable in that it circumvents the parallel implementation of the two filters. However we should thus choose $r_2(\cdot)$ to approximate the predictor. This will be illustrated in Section 3.3.2. We will also discuss in Section 4 how subsampling can be used to reduce the computational cost of calculating $\hat{\lambda}(\theta)$ to $\mathcal{O}(K)$.

3.1.2 Non-Canonical Case

In the case of latent-state estimation under the non-canonical case, the unknown $\lambda(\theta)$ inherent in (3.2) is replaced by $\hat{\lambda}(\theta)$, that is we divide the R.H.S. of (3.2) and (3.3) to obtain the estimator

$$\tilde{\psi}_U = \sum_{\mathbf{k} \in \mathbb{Z}^M} L(\mathbf{X}_U^{\mathbf{k}}) \psi(\mathbf{X}_U^{\mathbf{k}}) H_U^{\mathbf{k}} / \sum_{\mathbf{k} \in \mathbb{Z}^M} L(\mathbf{X}_U^{\mathbf{k}}) H_U^{\mathbf{k}}. \quad (3.5)$$

We can rewrite the above estimate in a standard form seen in the literature, and reduce the cost of computation to $\mathcal{O}(K)$. For example, if there is only one particle filter and we select w_t to satisfy (2.2), then the estimate reduces to $K^{-1} \sum_{k=1}^K \psi(\mathbf{X}_U^k)$, which is the standard estimate in the literature.

3.2 Unbiasedness and CLT under the canonical case

Let $f_0 = \psi_U$ and define,

$$f_t(\mathbf{x}_t) = E_q[\psi(\mathbf{X}_U) L(\mathbf{X}_U) | \mathbf{X}_t = \mathbf{x}_t], \quad 1 \leq t \leq U, \quad (3.6)$$

where E_q denotes expectation with respect to the importance densities q_t as defined in (2.7). There is no resampling involved under E_q . Let $\#_t^k$ denotes the number of copies of $\tilde{\mathbf{X}}_t^k$ generated from $(\tilde{\mathbf{X}}_t^1, \dots, \tilde{\mathbf{X}}_t^K)$ to form $(\mathbf{X}_t^1, \dots, \mathbf{X}_t^K)$. Thus conditionally, $(\#_t^1, \dots, \#_t^K) \sim$

Multinomial (K, W_t^1, \dots, W_t^K) . Let \mathcal{F}_{2t-1} and \mathcal{F}_{2t} denote the σ -algebras generated by all random variables just before and just after resampling respectively, at the t^{th} stage. In the case of $(m-1)T < t \leq mT$ for $m > 1$, these σ -algebras include all random variables generated by PF1 to PF $(m-1)$. Let E_K denote expectation with respect to K sample paths generated in each particle filter.

Theorem 1. *Define, for $(m-1)T < t \leq mT$ and $1 \leq j \leq K$,*

$$\begin{aligned}\epsilon_{2t-1}^j &= K^{-m+1} \sum_{\mathbf{k} \in Z^m: A_{m,t-1}^{k(m)}=j} [f_t(\tilde{\mathbf{X}}_t^{\mathbf{k}}) - f_{t-1}(\mathbf{X}_{t-1}^{\mathbf{k}})] H_{t-1}^{\mathbf{k}}, \\ \epsilon_{2t}^j &= K^{-m+1} \sum_{\mathbf{k} \in Z^m: A_{m,t-1}^{k(m)}=j} (\#_t^{k(m)} - KW_t^{k(m)}) f_t(\mathbf{X}_t^{\mathbf{k}}) \tilde{H}_t^{\mathbf{k}}.\end{aligned}\tag{3.7}$$

Then for each j and m , $\{\epsilon_u^j, \mathcal{F}_u, 2(m-1)T < u \leq 2mT\}$ is a martingale difference sequence, and

$$K(\hat{\psi}_U - \psi_U) = \sum_{m=1}^M \sum_{j=1}^K (\epsilon_{2(m-1)T+1}^j + \dots + \epsilon_{2mT}^j).\tag{3.8}$$

Therefore $E_K \hat{\psi}_U = \psi_U$.

Proof. Since $\#_t^k \sim \text{Binomial}(K, W_t^k)$ when conditioned on \mathcal{F}_{2t-1} , by the tower law of conditional expectations,

$$\begin{aligned}E_K(\epsilon_{2t-1}^j | \mathcal{F}_{2t-2}) &= K^{-m+1} \sum_{\mathbf{k} \in Z^m: A_{t-1}^{k(m)}=j} \{E_K[f_t(\tilde{\mathbf{X}}_t^{\mathbf{k}} | \mathcal{F}_{2t-2}) - f_{t-1}(\mathbf{X}_{t-1}^{\mathbf{k}})]\} H_{t-1}^{\mathbf{k}} = 0, \\ E_K(\epsilon_{2t}^j | \mathcal{F}_{2t-1}) &= K^{-m+1} \sum_{\mathbf{k} \in Z^m: A_{t-1}^{k(m)}=j} [E_K(\#_t^{k(m)} | \mathcal{F}_{2t-1}) - KW_t^{k(m)}] f_t(\tilde{\mathbf{X}}_t^{\mathbf{k}}) \tilde{H}_t^{\mathbf{k}} = 0,\end{aligned}$$

therefore $\{\epsilon_u^j, \mathcal{F}_u, 2(m-1)T < u \leq 2mT\}$ are indeed martingale difference sequences.

It follows from (2.5) and (2.10) that

$$\begin{aligned}& \sum_{\mathbf{k} \in Z^m: A_{t-1}^{k(m)}=j} (\#_t^{k(m)} - KW_t^{k(m)}) f_t(\tilde{\mathbf{X}}_t^{\mathbf{k}}) \tilde{H}_t^{\mathbf{k}} \\ &= \sum_{\mathbf{k} \in Z^m: A_t^{k(m)}=j} f_t(\mathbf{X}_t^{\mathbf{k}}) H_t^{\mathbf{k}} - \sum_{\mathbf{k} \in Z^m: A_{t-1}^{k(m)}=j} f_t(\tilde{\mathbf{X}}_t^{\mathbf{k}}) H_{t-1}^{\mathbf{k}},\end{aligned}$$

therefore by (3.7) and the cancellation of terms in a telescoping series,

$$\begin{aligned}\sum_{u=2(m-1)T+1}^{2mT} \epsilon_u^j &= K^{-m+1} \sum_{\mathbf{k} \in Z^m: A_{m,mT}^{k(m)}=j} f_{mT}(\mathbf{X}_{mT}^{\mathbf{k}}) H_{mT}^{\mathbf{k}} \\ &\quad - K^{-m+1} \sum_{\mathbf{k} \in Z^{m-1}} f_{(m-1)T}(\mathbf{X}_{(m-1)T}^{\mathbf{k}}) H_{(m-1)T}^{\mathbf{k}}.\end{aligned}\tag{3.9}$$

Therefore

$$\begin{aligned} \sum_{j=1}^K \left(\sum_{u=2(m-1)T+1}^{2mT} \epsilon_u^j \right) &= K^{-m+1} \sum_{\mathbf{k} \in \mathbb{Z}^m} f_{mT}(\mathbf{X}_{mT}^{\mathbf{k}}) H_{mT}^{\mathbf{k}} \\ &\quad - K^{-m+2} \sum_{\mathbf{k} \in \mathbb{Z}^{m-1}} f_{(m-1)T}(\mathbf{X}_{(m-1)T}^{\mathbf{k}}) H_{(m-1)T}^{\mathbf{k}}. \end{aligned} \quad (3.10)$$

By (3.2), the identity (3.8) follows from adding (3.10) over $1 \leq m \leq M$, keeping in mind that $f_0 = \psi_U$ and $f_U(\mathbf{x}_U) = L(\mathbf{x}_U)\psi(\mathbf{X}_U)$. \square

The martingale difference expansion (3.8) is for the purpose of standard error estimation, see Section 3.3.3. An alternative expansion, for the purpose of CLT theory in the spirit of [10, Chapter 8], is formed from the martingale difference sequence $\{(Z_u^1, \dots, Z_u^K) : 2(m-1)T < u \leq 2mT\}$, where

$$\begin{aligned} Z_{2t-1}^k &= K^{-m+1} \sum_{\mathbf{k} \in \mathbb{Z}^m : k(m)=k} [f_t(\tilde{\mathbf{X}}_t^{\mathbf{k}}) - f_{t-1}(\mathbf{X}_{t-1}^{\mathbf{k}})] H_{t-1}^{\mathbf{k}}, \\ Z_{2t}^k &= K^{-m+1} \sum_{\mathbf{k} \in \mathbb{Z}^m : k(m)=k} f_t(\mathbf{X}_t^{\mathbf{k}}) H_t^{\mathbf{k}} - K^{-m+1} \sum_{\mathbf{k} \in \mathbb{Z}^m} W_t^k f_t(\tilde{\mathbf{X}}_t^{\mathbf{k}}) \tilde{H}_t^{\mathbf{k}}. \end{aligned} \quad (3.11)$$

Analogous to (3.8), we have the expansion

$$K(\hat{\psi}_U - \psi_U) = \sum_{u=1}^{2U} (Z_u^1 + \dots + Z_u^K), \quad (3.12)$$

from which we can also conclude that $\hat{\psi}_U$ is unbiased.

The technical difficulties in working with (3.11) to prove the CLT is considerably more involved compared to the standard single particle filter, as there is now a sum over a multi-dimensional space. Therefore let us provide some intuitions first, focusing on the key arguments in the extension of the CLT to $M = 2$ segments.

For $t > T$, let

$$f_{2,t}(\mathbf{x}_{2,t}) = E_q[f_t(\mathbf{X}_t) | \mathbf{X}_{2,t} = \mathbf{x}_{2,t}].$$

By a ‘law of large numbers’ argument, see Lemma 2 in Appendix A.1,

$$\begin{aligned} K^{-1} \sum_{k=1}^K f_t(\tilde{\mathbf{X}}_t^{k\ell}) H_T^k &\doteq f_{2,t}(\tilde{\mathbf{X}}_{2,t}^\ell), \\ K^{-1} \sum_{k=1}^K f_t(\mathbf{X}_t^{k\ell}) H_T^k &\doteq f_{2,t}(\mathbf{X}_{2,t}^\ell). \end{aligned} \quad (3.13)$$

Therefore by (2.10) and (3.11), $Z_u^\ell \doteq Z_{2,u}^\ell$ for $2T < u \leq 2U$, where

$$\begin{aligned} Z_{2,2t-1}^\ell &= [f_{2,t}(\tilde{\mathbf{X}}_{2,t}^\ell) - f_{2,t-1}(\mathbf{X}_{2,t-1}^\ell)]H_{2,t-1}^\ell, \\ Z_{2,2t}^\ell &= f_{2,t}(\mathbf{X}_{2,t}^\ell)H_{2,t}^\ell - \sum_{j=1}^K W_t^j f_{2,t}(\tilde{\mathbf{X}}_{2,t}^j) \tilde{H}_{2,t}^j. \end{aligned} \quad (3.14)$$

We now have a martingale difference sequence $\{(Z_u^1, \dots, Z_u^K), \mathcal{F}_u, 1 \leq u \leq 2T\}$ that depends on the outcomes of PF1 only, and another sequence $\{(Z_{2,u}^1, \dots, Z_{2,u}^K), \mathcal{G}_u, 2T < u \leq 2U\}$ that depends on the outcomes of PF2 only, where \mathcal{G}_{2t-1} and \mathcal{G}_{2t} denote the σ -algebras generated by random variables in PF2 only, just before and just after resampling respectively, at stage t . Moreover,

$$K(\hat{\psi}_U - \psi_U) \doteq \sum_{u=1}^{2T} \left(\sum_{k=1}^K Z_u^k \right) + \sum_{u=2T+1}^{2U} \left(\sum_{\ell=1}^K Z_{2,u}^\ell \right). \quad (3.15)$$

Therefore subject to negligible error in (3.15), $\sqrt{K}(\hat{\psi}_U - \psi_U)$ is asymptotically normal, with variance the sum of the variance components due to each particle filter.

More generally in the case of M independent particle filters, define

$$f_{m,t}(\mathbf{x}_{m,t}) = E_q[f_t(\mathbf{X}_t) | \mathbf{X}_{m,t} = \mathbf{x}_{m,t}], \quad (3.16)$$

and recall the definition of $h_{m,t}$ in (2.8).

Theorem 2. *Let $\sigma^2 = \sum_{u=1}^{2U} \sigma_u^2$, where for $(m-1)T < t \leq mT$,*

$$\begin{aligned} \sigma_{2t-1}^2 &= E_q\{[f_{m,t}^2(\mathbf{X}_{m,t}) - f_{m,t-1}^2(\mathbf{X}_{m,t-1})]h_{m,t-1}(\mathbf{X}_{m,t-1})\}, \\ \sigma_{2t}^2 &= E_q\{[f_{m,t}(\mathbf{X}_{m,t})h_{m,t}(\mathbf{X}_{m,t}) - f_0]^2/h_{m,t}(\mathbf{X}_{m,t})\}. \end{aligned} \quad (3.17)$$

Assume that $E_q\{f_t^2(\mathbf{X}_t)[h_t(\mathbf{X}_t) + h_{t-1}(\mathbf{X}_t)]\} < \infty$ for $1 \leq t \leq U$. Then $\sigma^2 < \infty$ and

$$\sqrt{K}(\hat{\psi}_U - \psi_U) \Rightarrow N(0, \sigma^2) \text{ as } K \rightarrow \infty. \quad (3.18)$$

3.3 Asymptotic theory in the non-canonical case

In the non-canonical case, the estimator $\tilde{\psi}_U$, see (3.5), can be approximated by $\hat{\psi}_U^c$, an unbiased estimator under the canonical case of the centered function

$$\psi^c(\mathbf{x}_U) := \psi(\mathbf{x}_U) - \psi_U.$$

Therefore, analogous to (3.6) and (3.16), we define $f_{m,(m-1)T}^c(\mathbf{x}_t) = 0$ and

$$f_{m,t}^c(\mathbf{x}_t) = E_q[\psi^c(\mathbf{X}_U)L(\mathbf{X}_U)|\mathbf{X}_{m,t} = \mathbf{x}_{m,t}], \quad (m-1)T < t \leq mT. \quad (3.19)$$

The corollary below then follows from Theorem 2.

Corollary 1. *Let $\sigma_c^2 = \sum_{u=1}^{2U} \sigma_{c,u}^2$, where for $(m-1)T < t \leq mT$,*

$$\begin{aligned} \sigma_{c,2t-1}^2 &= E_q\{[f_{m,t}^c(\mathbf{X}_{m,t})]^2 - [f_{m,t-1}^c(\mathbf{X}_{m,t-1})]^2\}h_{m,t-1}(\mathbf{X}_{m,t-1}), \\ \sigma_{c,2t}^2 &= E_q\{[f_{m,t}^c(\mathbf{X}_{m,t})]^2 h_{m,t}(\mathbf{X}_{m,t})\}. \end{aligned} \quad (3.20)$$

Under the assumptions of Theorem 2,

$$\sqrt{K}(\tilde{\psi}_U - \psi_U) \Rightarrow N(0, \sigma_c^2) \text{ as } K \rightarrow \infty. \quad (3.21)$$

In Section 3.3.3, we show how σ_c^2 can be estimated in-sample, and discuss the implications in particle size allocation. Before that, we shall illustrate, in Sections 3.3.1 and 3.3.2, the advantage of segmentation in providing stability to smoothed latent-state estimations, that is estimation of $E_p(X_u|\mathbf{Y}_U)$ for $u < U$, with u fixed as $U \rightarrow \infty$.

3.3.1 Example

Consider the linear time-series

$$X_t = aX_{t-1} + \epsilon_t, \quad Y_t = X_t + \eta_t, \quad (3.22)$$

with $0 < a < 1$, $\epsilon_t \sim N(0, (1-a^2)\sigma_X^2)$ and $\eta_t \sim N(0, \sigma_Y^2)$. Let $\theta = (a, \sigma_X^2, \sigma_Y^2)$.

We shall illustrate on this simple example the advantage of parallel particle filters in smoothed estimation of X_t . Consider firstly the segmented particle filter with $T = 1$ and $q_t(x_t|\mathbf{x}_{t-1}) = p_\theta(x_t|\mathbf{Y}_U)$. Let $w_t \equiv 1$, therefore $h_{t,t} \equiv 1$ (recall also that $h_{t,t-1} \equiv 1$) for all t . Consider $\psi(\mathbf{x}_U) = x_u$ for some $1 \leq u < U$. By (3.1) and (3.19),

$$f_{t,t}^c(\mathbf{x}_t) = E_p(X_u|X_t = x_t, \mathbf{Y}_U) - E_p(X_u|\mathbf{Y}_U).$$

Since $f_{t,t-1}^c \equiv 0$, therefore by (3.20),

$$\sigma_{c,2t}^2 = \sigma_{c,2t-1}^2 = \text{Var}_p(E_p(X_u|X_t, \mathbf{Y}_U)|\mathbf{Y}_U) \leq \text{Var}_p(E_p(X_u|X_t)) = a^{|u-t|}\sigma_X^2, \quad (3.23)$$

with the inequality in (3.23) following from say eq.(2.5.3) of [1]. Therefore

$$\sigma_c^2 = \sum_{t=1}^U (\sigma_{c,2t-1}^2 + \sigma_{c,2t}^2) \leq \left(\frac{4}{1-a} - 2 \right) \sigma_X^2.$$

Corollary 1 says that, in this example, σ_c^2 is uniformly bounded and hence the estimation of X_u is stable as $U \rightarrow \infty$.

For the standard particle filter with no latent-state sequence segmentation, that is in the case $M = 1$ and $T = U$, we consider $q_t(x_t|\mathbf{x}_{t-1}) = p_\theta(x_t|\mathbf{x}_{t-1}, \mathbf{Y}_U)$ and $w_t \equiv 1$. Then by (3.19),

$$\begin{aligned} f_{2t}^c(\mathbf{x}_t)[= f_{1,2t}^c(\mathbf{x}_t)] &= E_p(X_u|\mathbf{X}_t = \mathbf{x}_t, \mathbf{Y}_U) - E_p(X_u|\mathbf{Y}_U), \\ &= X_u - E_p(X_u|\mathbf{Y}_U) \text{ if } t \geq u. \end{aligned}$$

Since $\text{Var}_p(X_u|\mathbf{Y}_U) \geq \text{Var}_p(X_u|X_{u-1}, X_{u+1}, Y_u) > 0$, therefore in this example, $\sigma_{c,2t}^2$ is bounded away from 0 for $t \geq u$, and consequently $\sigma_c^2 \rightarrow \infty$ as $U \rightarrow \infty$.

Intuitively, in the case of the standard particle filter, the estimation of $E_p(X_u|\mathbf{Y}_U)$ is unstable as $U \rightarrow \infty$ simply due to degeneracy caused by repeated resamplings at $t \geq u$. On the other hand, the repeated resamplings does not cause instability in the segmented method because resampling in one segment does not result in sample depletion of another segment. There is a vast literature on smoothed latent-state estimators, see for example [5]. We do not go into details here as our main motivation for looking at parallel particle filters is to achieve wall-clock computation time savings; the variance reductions in smoothed estimates can be viewed as an added benefit.

3.3.2 Numerical Study

As in Section 3.3.1, consider the linear time-series (3.22) and the estimation of $\psi(\mathbf{x}_U) = x_u$ for $0 < u \leq U$, conditioned on \mathbf{Y}_U . Kalman updating formulas are applied to compute $E(X_u|\mathbf{Y}_{u-1})$ and $E(X_u|\mathbf{Y}_u)$ analytically, and the Rauch-Tung-Striebel smoother is applied to compute $E(X_u|\mathbf{Y}_t)$ for $t > u$. The mean-squared errors (MSE) of the particle filter methods can then be computed using Monte Carlo.

The first method we consider is the standard particle filter that performs bootstrap resampling at every step. We select q_1 as $N(0, \sigma_X^2)$ and $q_t(\cdot|\mathbf{x}_{t-1})$ as $N(ax_{t-1}, (1-a^2)\sigma_X^2)$

u	$E_p(X_u \mathbf{Y}_{50})$	MSE ($\times 10^{-2}$)		
		Standard	Seg.: $N(0, 1)$	Seg.: $N(\hat{\mu}_m, \hat{\sigma}_m^2)$
5	-0.46	5.8 \pm 1.0	0.67 \pm 0.09	0.6 \pm 0.1
10	0.45	6.0 \pm 0.8	0.35 \pm 0.04	0.34 \pm 0.05
15	0.80	5.1 \pm 0.6	0.64 \pm 0.08	0.58 \pm 0.07
20	0.95	4.9 \pm 0.7	1.1 \pm 0.2	0.77 \pm 0.09
25	1.16	6.1 \pm 1.0	2.8 \pm 0.4	3.0 \pm 0.4
30	2.81	2.7 \pm 0.4	2.2 \pm 0.3	1.0 \pm 0.1
35	0.81	1.8 \pm 0.2	1.9 \pm 0.2	0.8 \pm 0.1
40	-1.10	1.4 \pm 0.2	1.5 \pm 0.2	0.9 \pm 0.2
45	-0.50	0.7 \pm 0.1	0.69 \pm 0.09	0.55 \pm 0.07
50	0.38	0.19 \pm 0.02	0.18 \pm 0.02	0.20 \pm 0.03

Table 1: MSE ($\times 10^{-2}$) of $\hat{E}(X_u|\mathbf{Y}_{50})$ for: 1. the standard particle filter, 2. the segmented particle filter initialized at $N(0, 1)$, 3. the segmented particle filter initialized at $N(\hat{\mu}_m, \hat{\sigma}_m^2)$, with $(\hat{\mu}_m, \hat{\sigma}_m^2)$ estimated from past observations.

for $t > 1$, hence

$$w_t(\mathbf{x}_t) = \exp \left[-\frac{(Y_t - x_t)^2}{2\sigma_Y^2} \right] \quad t \geq 1. \quad (3.24)$$

We next consider parallel particle filters with \mathbf{X}_U segmented into M portions of equal length $T = U/M$. The resampling weights are as in (3.24), and like the standard particle filter,

$$q_{m,t}(\cdot|\mathbf{x}_{t-1}) = N(ax_{t-1}, (1-a^2)\sigma_X^2) \text{ for } (m-1)T+1 < t \leq mT.$$

We consider two versions of $q_{m,(m-1)T+1}$ ($= r_m$ here), the initial sampling distribution of $\tilde{X}_{m,(m-1)T+1}^k$. In the first version, we simply let $q_{m,(m-1)T+1} = N(0, \sigma_X^2)$. In the second version, we let $q_{m,(m-1)T+1} = N(\hat{\mu}_m, \hat{\sigma}_m^2)$, with $(\hat{\mu}_1, \hat{\sigma}_1^2) = (0, \sigma_X^2)$ and for $m \geq 2$, $(\hat{\mu}_m, \hat{\sigma}_m^2)$ simulated using $(Y_{(m-1)T-r}, \dots, Y_{(m-1)T})$ for some $r \geq 0$. More specifically, we let

$$\begin{aligned} \mu_m &= E(X_{(m-1)T+1} | Y_{(m-1)T-r}, \dots, Y_{(m-1)T}), \\ \sigma_m^2 &= \text{Var}(X_{(m-1)T+1} | Y_{(m-1)T-r}, \dots, Y_{(m-1)T}), \end{aligned}$$

and estimate them by sampling $(X_{(m-1)T-r}, \dots, X_{(m-1)T+1})$ using particle filters.

In our simulation study, we select $a = 0.8$, $\sigma_X^2 = \sigma_Y^2 = 1$, $U = 50$ and $M = 5$. We apply $K = 500$ particles in each filter, and for the second version of the segmented method, we consider $r = 4$. Each method is repeated 100 times, and its MSE are reported in Table 1. We see from Table 1 substantial variance reductions for parallel particle filters over standard particle filters, especially when $U - u$ is big, agreeing with the discussions in Section 3.3.1. In addition, we see that applying estimation of (μ_m, σ_m^2) improves upon the performances of parallel particle filters.

3.3.3 In-sample variance estimation and particle size allocation

Let $\sigma_{Pm}^2 = \sum_{u=2(m-1)T+1}^{2mT} \sigma_{c,u}^2$ be the variability attributed to the m th particle filter. If K_m particles are allocated to particle filter m with K_m large, then analogous to (3.21),

$$\text{Var}(\tilde{\psi}_U) \doteq \sum_{m=1}^M \frac{\sigma_{Pm}^2}{K_m}. \quad (3.25)$$

Therefore being able to estimate σ_{Pm}^2 in-sample would allow us to optimally allocate the particle sizes in the M particle filters so as to minimize (3.25). The estimation can be done in the following manner.

Consider $1 \leq m \leq M$ and let $C_m^j = \{k : A_{m,mT}^k = j\}$, noting that $A_{m,mT}^k = j$ if and only if $\mathbf{X}_{m,mT}^k$ is descended from $\tilde{X}_{(m-1)T+1}^j$. Let

$$Q_j(\mu) = \sum_{\mathbf{k} \in \mathbb{Z}^M : k(m) \in C_m^j} L(\mathbf{X}_U^{\mathbf{k}}) [\psi(\mathbf{X}_U^{\mathbf{k}}) - \mu] H_U^{\mathbf{k}}.$$

Theorem 3. *Under the assumptions of Corollary 1,*

$$\hat{\sigma}_{Pm}^2(\psi_U) := K^{-1} \sum_{j=1}^K Q_j^2(\psi_U) \xrightarrow{P} \sigma_{Pm}^2, \quad 1 \leq m \leq M. \quad (3.26)$$

Since $\tilde{\psi}_U \xrightarrow{P} \psi_U$ by Corollary 1, therefore $\hat{\sigma}_{Pm}^2(\tilde{\psi}_U)$ is consistent for σ_{Pm}^2 . Besides particle size allocation, being able to estimate σ_{Pm}^2 , $1 \leq m \leq M$, and hence σ_c^2 allows us to assess the level of accuracy of $\tilde{\psi}_U$ in estimating ψ_U .

4 Discussion

We discuss here the subsampling approach, proposed in [4, 13], that can be used to reduce the $\mathcal{O}(K^2)$ computational cost of our estimates. We make the discussions more concrete here by considering $M = 2$ and focusing on the estimation of the likelihood $\lambda(\theta)$.

The actual computational cost of the double sum in (3.4) may be less expensive than it seems, given that this operation is done only once, and that time-savings can be achieved if we bother to first group the segments $\mathbf{X}_{2,U}^\ell$ having a common first component. However asymptotically, we do have a larger computational complexity due to the double sum.

Let $\{(k(v), \ell(v)) : 1 \leq v \leq V\}$ be selected i.i.d. from β , a positive distribution on Z_K^2 , and estimate $\lambda(\theta)$ by

$$\hat{\lambda}^*(\theta) = \left(\prod_{t=1}^U \bar{w}_t \right) (K^2 V)^{-1} \sum_{v=1}^V \frac{p_\theta(X_{T+1}^{\ell(v)} | X_T^{k(v)})}{r_2(X_{T+1}^{\ell(v)}) \beta(k(v), \ell(v))}.$$

Since $\hat{\lambda}(\theta)$ is unbiased for $\lambda(\theta)$, therefore so is $\hat{\lambda}^*(\theta)$. For example, we can apply stratification sampling so that “good” pairs are chosen more frequently. The choice of $V \sim K^s$ for $s = 1$ would give us a $\mathcal{O}(K)$ algorithm, though we may have to select $s > 1$ in order to maintain the asymptotic variance of $\hat{\lambda}(\theta)$. As the computation of $\hat{\lambda}^*(\theta)$ is separate from the execution of the parallel particle filters and can be done off-line, improving $\hat{\lambda}^*(\theta)$ with more sampling does not require re-running of the particle filters.

Acknowledgements

The third author was supported by an MOE Singapore grant R-155-000-119-133 and is affiliated with the Risk Management Institute and Centre for Quantitative Finance at NUS.

A Proofs

We preface the proofs of the main results of Section 3 with two supporting lemmas in Appendix A.1 below. Lemma 1 is a weak law of large number for sums of segmented sequences. Lemma 2 provides finer approximations of such sums.

A.1 Asymptotics and finer approximations for sums of segmented sequences

Lemma 1. *Let G be a real-valued measurable function of \mathbf{x}_t for some $(m-1)T < t \leq mT$ with $1 \leq m \leq M$.*

(a) *If $\tilde{\mu}_t := E_q[G(\mathbf{X}_t)/h_{t-1}(\mathbf{X}_{t-1})]$ exists and is finite, then as $K \rightarrow \infty$,*

$$K^{-m} \sum_{\mathbf{k} \in \mathbb{Z}^m} G(\tilde{\mathbf{X}}_t^{\mathbf{k}}) \xrightarrow{p} \tilde{\mu}_t.$$

(b) *If $\mu_t := E_q[G(\mathbf{X}_t)/h_t(\mathbf{X}_t)]$ exists and is finite, then as $K \rightarrow \infty$,*

$$K^{-m} \sum_{\mathbf{k} \in \mathbb{Z}^m} G(\mathbf{X}_t^{\mathbf{k}}) \xrightarrow{p} \mu_t.$$

(c) *For each $\mathbf{k} \in \mathbb{Z}^m$,*

$$\frac{\tilde{H}_t^{\mathbf{k}}}{h_t(\tilde{\mathbf{X}}_t^{\mathbf{k}})} = \frac{H_t^{\mathbf{k}}}{h_t(\mathbf{X}_t^{\mathbf{k}})} = \frac{\bar{w}_1 \cdots \bar{w}_t}{\eta_t} \xrightarrow{p} 1. \quad (\text{A.1})$$

Proof. Since $G = G^+ - G^-$, we can assume without loss of generality that G is nonnegative. The proofs of (a) and (b) for $t \leq T$ follows from standard induction arguments, see [6, Lemma 2]. For $t > T$, induction arguments are again used, but the framework is now considerably more complicated with summation on a multi-dimensional instead of a one-dimensional space. Unlike in [6], characteristic functions are now needed in the induction proof.

Let $T < u \leq 2T$ and assume that $\tilde{\mu}_u$ exists and is finite, and that Lemma 1(b) holds for $t = u - 1$. Let $V_u^\ell = K^{-1} \sum_{k=1}^K G(\tilde{\mathbf{X}}_u^{k\ell})$ and consider the decomposition $V_u^\ell = R_{u,c}^\ell + S_{u,c}^\ell$, where

$$R_{u,c}^\ell = K^{-1} \sum_{k=1}^K G(\tilde{\mathbf{X}}_u^{k\ell}) \mathbf{1}_{\{G(\tilde{\mathbf{X}}_u^{k\ell}) \leq c\}}, \quad S_{u,c}^\ell = K^{-1} \sum_{k=1}^K G(\tilde{\mathbf{X}}_u^{k\ell}) \mathbf{1}_{\{G(\tilde{\mathbf{X}}_u^{k\ell}) > c\}}. \quad (\text{A.2})$$

Let $\bar{V}_u = K^{-1} \sum_{\ell=1}^K V_u^\ell$, and define $\bar{R}_{u,c}, \bar{S}_{u,c}$ in a similar fashion. Let $\mathbf{i} = \sqrt{-1}$ and define

$$\begin{aligned} \tilde{\mu}_{u,c} &= E_q[G(\mathbf{X}_u) \mathbf{1}_{\{G(\mathbf{X}_u) > c\}} / h_{u-1}(\mathbf{X}_{u-1})], \\ \varphi_{2u-1,c}^\ell(\theta | \mathcal{F}_{2u-2}) &= E_K[\exp(\mathbf{i}\theta K^{-1} R_{u,c}^\ell) | \mathcal{F}_{2u-2}]. \end{aligned} \quad (\text{A.3})$$

Let $\delta > 0$. Since $R_{u,\delta K}^1, \dots, R_{u,\delta K}^K$ are independent conditioned on \mathcal{F}_{2u-2} ,

$$\begin{aligned} E_K[\exp(\mathbf{i}\theta \bar{V}_u) | \mathcal{F}_{2u-2}] &= \prod_{\ell=1}^K \varphi_{2u-1,\delta K}^\ell(\theta | \mathcal{F}_{2u-2}) + r_K, \\ r_K &= E_K\{\exp(\mathbf{i}\theta \bar{R}_{u,\delta K})[\exp(\mathbf{i}\theta \bar{S}_{u,\delta K}) - 1] | \mathcal{F}_{2u-2}\}. \end{aligned} \quad (\text{A.4})$$

Since $|e^{\mathbf{i}z} - 1| \leq |z|$, for $K \geq c/\delta$,

$$|r_K| \leq |\theta| K^{-1} \sum_{\ell=1}^K E_K(S_{u,c}^\ell | \mathcal{F}_{2u-2}). \quad (\text{A.5})$$

By the induction hypothesis applied on $G_c(\mathbf{x}_{u-1}) := E_q[G(\mathbf{X}_u) \mathbf{1}_{\{G(\mathbf{x}_u) > c\}} | \mathbf{X}_{u-1} = \mathbf{x}_{u-1}]$,

$$K^{-1} \sum_{\ell=1}^K E_K(S_{u,c}^\ell | \mathcal{F}_{2u-2}) \xrightarrow{P} E_q[G_c(\mathbf{X}_{u-1}) / h_{u-1}(\mathbf{X}_{u-1})] (= \tilde{\mu}_{u,c}). \quad (\text{A.6})$$

Since $\tilde{\mu}_{u,c} \rightarrow 0$ as $c \rightarrow \infty$, therefore by (A.5) and (A.6),

$$r_K \xrightarrow{P} 0. \quad (\text{A.7})$$

Let $\tilde{R}_{u,c} = R_{u,c}^\ell - E_K(R_{u,c}^\ell | \mathcal{F}_{2u-2})$ and

$$\tilde{\varphi}_{2u-1,c}^\ell(\theta | \mathcal{F}_{2u-2}) = E_K[\exp(\mathbf{i}\theta K^{-1} \tilde{R}_{u,c}^\ell) | \mathcal{F}_{2u-2}]. \quad (\text{A.8})$$

Since $|E[e^{\mathbf{i}\theta Z} - (1 + \mathbf{i}\theta Z - \frac{\theta^2 Z^2}{2})]| \leq \theta^2 E Z^2$ (see (26.5) of [3]) and $(R_{u,\delta K}^\ell)^2 \leq \delta K R_{u,\delta K}^\ell$,

$$\begin{aligned} &|\tilde{\varphi}_{2u-1,\delta K}^\ell(\theta | \mathcal{F}_{2u-2}) - \{1 - [\theta^2 / (2K^2)] \text{Var}_K(R_{u,\delta K}^\ell | \mathcal{F}_{2u-2})\}| \\ &\leq (\theta/K)^2 \text{Var}_K(R_{u,\delta K}^\ell | \mathcal{F}_{2u-2}) \leq (\theta/K)^2 E_K[(R_{u,\delta K}^\ell)^2 | \mathcal{F}_{2u-2}] \\ &\leq (\delta \theta^2 / K) E_K(R_{u,\delta K}^\ell | \mathcal{F}_{2u-2}). \end{aligned} \quad (\text{A.9})$$

Since $|\prod_{\ell=1}^K z_\ell - \prod_{\ell=1}^K y_\ell| \leq \sum_{\ell=1}^K |z_\ell - y_\ell|$ whenever $|z_\ell| \leq 1$ and $|y_\ell| \leq 1$ for all ℓ (see Lemma 1 in Section 2.7 of [3]), by the induction hypothesis applied on $G'_c(\mathbf{x}_u) = E_q[G(\mathbf{X}_u) \mathbf{1}_{\{G(\mathbf{x}_u) \leq c\}} | \mathbf{X}_{u-1} = \mathbf{x}_{u-1}]$ and then letting $c \rightarrow \infty$,

$$\begin{aligned} &\left| \prod_{\ell=1}^K \tilde{\varphi}_{2u-1,\delta K}^\ell(\theta | \mathcal{F}_{2u-2}) - \prod_{\ell=1}^K \{1 - [\theta^2 / (2K^2)] \text{Var}_K(R_{u,\delta K}^\ell | \mathcal{F}_{2u-2})\} \right| \\ &\leq \delta \theta^2 K^{-1} \sum_{\ell=1}^K E_K(R_{u,\delta K}^\ell | \mathcal{F}_{2u-2}) \xrightarrow{P} \delta \theta^2 \tilde{\mu}_u. \end{aligned} \quad (\text{A.10})$$

Let $\delta_0 > 0$ be such that $\log(1-y) \geq -2y$ for $0 < y < (\theta\delta_0)^2$. Therefore by the inequalities in (A.9), for $0 < \delta \leq \delta_0$,

$$\begin{aligned} & \prod_{\ell=1}^K \{1 - [\theta^2/(2K^2)] \text{Var}_K(R_{u,\delta K}^\ell | \mathcal{F}_{2u-2})\} \\ & \geq \prod_{\ell=1}^K \{1 - [\delta\theta^2/(2K)] E_K[(R_{u,\delta K}^\ell)^2 | \mathcal{F}_{2u-2}]\} \\ & \geq \exp \left[-\delta\theta^2 K^{-1} \sum_{\ell=1}^K E_K(R_{u,\delta K}^\ell | \mathcal{F}_{2u-2}) \right] \xrightarrow{P} \exp(-\delta\theta^2 \tilde{\mu}_u). \end{aligned} \quad (\text{A.11})$$

By the definitions of $\varphi_{2u-1,c}^\ell$ and $\tilde{\varphi}_{2u-1,c}^\ell$ in (A.3) and (A.8),

$$\begin{aligned} & \prod_{\ell=1}^K [\varphi_{2u-1,\delta K}^\ell(\theta | \mathcal{F}_{2u-2}) / \tilde{\varphi}_{2u-1,\delta K}^\ell(\theta | \mathcal{F}_{2u-2})] \\ & = \exp \left[\mathbf{i}\theta K^{-1} \sum_{\ell=1}^K E_K(R_{u,\delta K}^\ell | \mathcal{F}_{2u-1}) \right] \xrightarrow{P} \exp(\mathbf{i}\theta \tilde{\mu}_u). \end{aligned} \quad (\text{A.12})$$

It follows from (A.4), (A.7) and (A.10)–(A.12), with $\delta \rightarrow 0$, that

$$E_K[\exp(\mathbf{i}\theta \bar{V}_u) | \mathcal{F}_{2u-2}] \xrightarrow{P} \exp(\mathbf{i}\theta \tilde{\mu}_u).$$

Therefore $E_K \exp(\mathbf{i}\theta \bar{V}_u) \xrightarrow{P} \exp(\mathbf{i}\theta \tilde{\mu}_u)$, equivalently $\bar{V}_u \xrightarrow{P} \tilde{\mu}_u$. Hence (a) holds for $t = u$ whenever (b) holds for $t = u-1$. By similar arguments, (b) holds for $t = u$ whenever (a) holds for $t = u$. The induction arguments to show (a) and (b) for $T < t \leq 2T$ are now complete. Similar induction arguments can be used to show (a) and (b) for $(m-1)T < t \leq mT$ for $m = 3, \dots, M$.

The identities in (A.1) follows from multiplying (2.9) over “ (m, t) ” = $(1, T), \dots, (m-1, (m-1)T), (m, t)$. By (2.6) and (2.8), applying (a) on $G = w_t$ gives us $\bar{w}_t \xrightarrow{P} \eta_t/\eta_{t-1}$. Therefore (c) holds. \square

Lemma 2. Let G_u be a measurable function of \mathbf{x}_u with $(m-1)T < u \leq mT$ and define $G_{m,u}(\mathbf{x}_{m,u}) = E_q[G_u(\mathbf{X}_u) | \mathbf{X}_{m,u} = \mathbf{x}_{m,u}]$.

(a) If $E_q[G_u^2(\mathbf{X}_u)h_{u-1}(\mathbf{X}_{u-1})] < \infty$, then

$$K^{-1} \sum_{\ell=1}^K \left[K^{-M+1} \sum_{\mathbf{k} \in \mathbb{Z}^{m-1}} G_u(\tilde{\mathbf{X}}_u^{\mathbf{k}\ell}) H_{u-1}^{\mathbf{k}\ell} - G_{m,u}(\tilde{\mathbf{X}}_{m,u}^\ell) H_{m,u-1}^\ell \right]^2 \xrightarrow{P} 0, \quad (\text{A.13})$$

$$\sum_{\ell=1}^K \left[K^{-M+1} \sum_{\mathbf{k} \in \mathbb{Z}^{m-1}} G_u(\tilde{\mathbf{X}}_u^{\mathbf{k}\ell}) H_{u-1}^{\mathbf{k}\ell} - G_{m,u}(\tilde{\mathbf{X}}_{m,u}^\ell) H_{m,u-1}^\ell \right] = o_p(K^{\frac{1}{2}}), \quad (\text{A.14})$$

where $\mathbf{k}\ell = (k(1), \dots, k(m-1), \ell)$.

(b) If $E_q[G_u^2(\mathbf{X}_u)h_u(\mathbf{X}_u)] < \infty$, then

$$K^{-1} \sum_{\ell=1}^K \left[K^{-M+1} \sum_{\mathbf{k} \in \mathbb{Z}^{m-1}} G_u(\mathbf{X}_u^{\mathbf{k}\ell}) H_u^{\mathbf{k}\ell} - G_{m,u}(\mathbf{X}_{m,u}^\ell) H_{m,u}^\ell \right]^2 \xrightarrow{p} 0, \quad (\text{A.15})$$

$$\sum_{\ell=1}^K \left[K^{-M+1} \sum_{\mathbf{k} \in \mathbb{Z}^{m-1}} G_u(\mathbf{X}_u^{\mathbf{k}\ell}) H_u^{\mathbf{k}\ell} - G_{m,u}(\mathbf{X}_{m,u}^\ell) H_{m,u}^\ell \right] = o_p(K^{\frac{1}{2}}). \quad (\text{A.16})$$

Proof. The case $m = 1$ is trivial, so let us first consider $m = 2$. For $1 \leq t \leq T$, let

$$G_{t,u}(\mathbf{x}_t, \mathbf{x}_{2,u}) = E_q[G_u(\mathbf{X}_u) | \mathbf{X}_t = \mathbf{x}_t, \mathbf{X}_{2,u} = \mathbf{x}_{2,u}].$$

By Lemma 1(c), $K^{-1} \sum_{k=1}^K G_u(\tilde{\mathbf{X}}_u^{k\ell}) H_{u-1}^{k\ell} - G_{2,u}(\tilde{\mathbf{X}}_{2,u}^\ell) H_{2,u-1}^\ell = [1 + o_p(1)] D^\ell$ uniformly over ℓ , where

$$D^\ell = \left[K^{-1} \sum_{k=1}^K G_u(\tilde{\mathbf{X}}_u^{k\ell}) H_T^{k\ell} - G_{2,u}(\tilde{\mathbf{X}}_{2,u}^\ell) \right] h_{2,u-1}(\mathbf{X}_{2,u-1}^\ell). \quad (\text{A.17})$$

By (A.17), we have the expansion

$$D^\ell = K^{-1} \sum_{s=1}^{2T} (d_s^{1\ell} + \dots + d_s^{K\ell}) h_{2,u-1}(\mathbf{X}_{2,u-1}^\ell), \quad (\text{A.18})$$

$$\text{where } d_{2t-1}^{k\ell} = [G_{t,u}(\tilde{\mathbf{X}}_t^k, \tilde{\mathbf{X}}_{2,u}^\ell) - G_{t-1,u}(\mathbf{X}_{t-1}^k, \tilde{\mathbf{X}}_{2,u}^\ell)] H_{t-1}^k,$$

$$d_{2t}^{k\ell} = G_{t,u}(\mathbf{X}_t^k, \tilde{\mathbf{X}}_{2,u}^\ell) H_t^k - \sum_{j=1}^K W_t^j G_{t,u}(\tilde{\mathbf{X}}_t^j, \tilde{\mathbf{X}}_{2,u}^\ell) H_{t-1}^k,$$

with the convention that for $t = 1$, $H_{t-1}^k = 1$ and $G_{t-1,u}(\mathbf{X}_{t-1}^k, \tilde{\mathbf{X}}_{2,u}^\ell) = G_{2,u}(\tilde{\mathbf{X}}_{2,u}^\ell)$.

Let $D_v^\ell = K^{-1} \sum_{s=1}^v (d_s^{1\ell} + \dots + d_s^{K\ell}) h_{2,u-1}(\mathbf{X}_{2,u-1}^\ell)$. We shall show inductively that uniformly over ℓ ,

$$\sum_{\ell=1}^K E_K[(D_v^\ell)^2 | \mathcal{G}_{2u-2}] = O_p(1), \quad v = 1, \dots, 2T, \quad (\text{A.19})$$

where \mathcal{G}_{2u-2} denotes the σ -algebra for all random variables generated in the second particle filter up to and including resampling at the $(u-1)$ th stage. Since $d_1^{1\ell}, \dots, d_1^{K\ell}$ are uncorrelated with mean 0 conditioned on \mathcal{G}_{2u-2} , by Lemma 1(a) and (c),

$$\sum_{\ell=1}^K E_K[(D_1^\ell)^2 | \mathcal{G}_{2u-2}] = K^{-2} \sum_{k=1}^K \sum_{\ell=1}^K E_K[(d_1^{k\ell})^2 | \mathcal{G}_{2u-2}] h_{2,u-1}(\mathbf{X}_{2,u-1}^\ell)^2 = O_p(1).$$

Therefore (A.19) holds for $v = 1$. Consider next $v > 1$. Let $\mathcal{H}_v = \mathcal{F}_v \cup \mathcal{G}_{2u-2}$. Since

$$D_v^\ell = D_{v-1}^\ell + K^{-1} (d_v^{1\ell} + \dots + d_v^{K\ell}) h_{2,u-1}(\mathbf{X}_{2,u-1}^\ell),$$

and $d_v^{1\ell}, \dots, d_v^{K\ell}$ are conditionally independent with mean 0 given \mathcal{H}_{v-1} , by Lemma 1(a) and (c),

$$\begin{aligned} \sum_{\ell=1}^K E_K[(D_v^\ell)^2 | \mathcal{H}_{v-1}] &= \sum_{\ell=1}^K (D_{v-1}^\ell)^2 + K^{-2} \sum_{k=1}^K \sum_{\ell=1}^K E_K[(d_v^{k\ell})^2 | \mathcal{H}_{v-1}] h_{2,u-1}^2(\mathbf{X}_{2,u-1}^\ell) \\ &= \sum_{\ell=1}^K (D_{v-1}^\ell)^2 + O_p(1). \end{aligned}$$

Therefore (A.19) for v follows from (A.19) for $v-1$. By induction, (A.19) holds for $1 \leq v \leq 2T$. In particular, since $D_{2T}^\ell = D^\ell$, (A.13) holds for $m = 2$.

By (A.19) for $v = 2T$, and noting that D^1, \dots, D^K are conditionally independent with mean 0 given \mathcal{F}_{2u-2} , and that $\mathcal{G}_{2u-2} \subset \mathcal{F}_{2u-2}$,

$$E_K \left[\left(\sum_{\ell=1}^K D^\ell \right)^2 \middle| \mathcal{F}_{2u-2} \right] = \sum_{\ell=1}^K E_K[(D^\ell)^2 | \mathcal{F}_{2u-2}] = O_p(K). \quad (\text{A.20})$$

Therefore $K^{-1}(\sum_{\ell=1}^K D^\ell)^2 \xrightarrow{p} 0$, and (A.14) holds for $m = 2$. The extension of the proof to $m > 2$ and the proofs of (A.15) and (A.16) apply similar arguments to those of (A.13) and (A.14). \square

A.2 Proofs of Theorem 2, Corollary 1 and Theorem 3

Proof of Theorem 2. Let $S = \sum_{m=1}^M S_m$, where

$$\begin{aligned} S_m &= \sum_{u=2(m-1)T+1}^{2mT} (Z_{m,u}^1 + \dots + Z_{m,u}^K), \\ Z_{m,2t-1}^k &= [f_{m,t}(\tilde{\mathbf{X}}_{m,t}^k) - f_{m,t-1}(\mathbf{X}_{m,t-1}^k)] H_{m,t}^k, \\ Z_{m,2t}^k &= f_{m,t}(\mathbf{X}_{m,t}^k) H_{m,t}^k - \sum_{j=1}^K W_t^j f_{m,t}(\tilde{\mathbf{X}}_{m,t}^j) \tilde{H}_{m,t}^j. \end{aligned}$$

By the CLT for particle filters on unsegmented HMM sequences, see for example [6],

$$K^{-\frac{1}{2}} S_m \Rightarrow N(0, \sigma_{\text{Pm}}^2), \quad 1 \leq m \leq M.$$

Since the particle filters operate independently and S_m depends only on the m th particle filter,

$$K^{-\frac{1}{2}} S \Rightarrow N(0, \sigma^2) \text{ where } \sigma^2 = \sum_{m=1}^M \sigma_{\text{Pm}}^2. \quad (\text{A.21})$$

By (3.12),

$$\sqrt{K}(\widehat{\psi}_U - \psi_U) = K^{-\frac{1}{2}}S + K^{-\frac{1}{2}} \sum_{u=1}^{2mU} \sum_{\ell=1}^K (Z_u^\ell - Z_{m,u}^\ell). \quad (\text{A.22})$$

Therefore by (3.11), (3.14) and (A.14) applied on $G_t(\mathbf{x}_t) = f_t(\mathbf{x}_t) - f_{t-1}(\mathbf{x}_{t-1})$,

$$K^{-\frac{1}{2}} \sum_{k=1}^K (Z_u^k - Z_{m,u}^k) \xrightarrow{P} 0 \quad (\text{A.23})$$

for $u = 2t - 1$. And by (3.11), (3.14) and (A.16) applied on $G_t(\mathbf{x}_t) = f_t(\mathbf{x}_t)$ and (A.16) applied on $G_t(\mathbf{x}_t) = w_t(\mathbf{x}_t)f_t(\mathbf{x}_t)$, (A.23) holds for $u = 2t$. We conclude Theorem 2 from (A.21)–(A.23). \square

Proof of Corollary 1. By (3.5),

$$\widetilde{\psi}_U - \psi_U = \left[K^{-M} \sum_{\mathbf{k} \in \mathbb{Z}^M} L(\mathbf{X}_U^{\mathbf{k}}) H(\mathbf{X}_U^{\mathbf{k}}) \right]^{-1} \widehat{\psi}_U^c, \quad (\text{A.24})$$

where $\widehat{\psi}_U^c = K^{-M} \sum_{\mathbf{k} \in \mathbb{Z}^M} L(\mathbf{X}_U^{\mathbf{k}}) \psi^c(\mathbf{X}_U^{\mathbf{k}}) H_U^{\mathbf{k}}$, is (3.2) with ψ replaced by $\psi^c(\mathbf{x}_U) = \psi(\mathbf{x}_U) - \psi_U$. By Theorem 2,

$$\sqrt{K} \widehat{\psi}_U^c \Rightarrow N(0, \sigma_c^2). \quad (\text{A.25})$$

By Lemma 1(b) and (c),

$$K^{-M} \sum_{\mathbf{k} \in \mathbb{Z}^M} L(\mathbf{X}_U^{\mathbf{k}}) H(\mathbf{X}_U^{\mathbf{k}}) \xrightarrow{P} 1,$$

and Corollary 1 therefore follows from (A.24) and (A.25). \square

Proof of Theorem 3. . We shall show Theorem 3 in detail for the case $M = m = 2$, assuming without loss of generality that $\psi_U = 0$. It follows from [6, Corollary 2] that

$$K^{-1} \sum_{j=1}^K (e^j)^2 \xrightarrow{P} \sigma_{\mathbb{P}^2}^2, \text{ where } e^j = \sum_{\ell: A_{2,U}^\ell = j} f_{2,U}(\mathbf{X}_{2,U}^\ell) H_{2,U}^\ell. \quad (\text{A.26})$$

By (3.11), (3.14) and (3.26),

$$\widehat{\sigma}_{\mathbb{P}^2}^2(0) = K^{-1} \sum_{j=1}^K \left(e^j + \sum_{v=2T+1}^{2U} \sum_{\ell: A_{2,t}^\ell = j} \zeta_v^\ell \right)^2,$$

where $t = \lfloor v/2 \rfloor$, with $\lfloor \cdot \rfloor$ denoting the greatest integer function,

$$\zeta_{2u-1}^\ell = Z_{2u-1}^\ell - Z_{2,2u-1}^\ell, \quad \zeta_{2u}^\ell = (\#_u^\ell - KW_u^\ell) \chi^\ell,$$

$$\chi^\ell = K^{-1} \sum_{k=1}^K f_u(\tilde{\mathbf{X}}^{k\ell}) \tilde{H}_u^{k\ell} - f_{2,u}(\mathbf{X}_{2,u}^\ell) \tilde{H}_{2,u}^\ell.$$

Therefore by (A.26), to show (3.26), it suffices to show that

$$K^{-1} \sum_{j=1}^K \left(\sum_{v=2T+1}^{2U} \sum_{\ell: A_{2,t}^\ell = j} \zeta_v^\ell \right)^2 \xrightarrow{P} 0. \quad (\text{A.27})$$

We shall apply induction to show that

$$K^{-1} \sum_{j=1}^K \left(\sum_{v=2T+1}^s \sum_{\ell: A_{2,t}^\ell = j} \zeta_v^\ell \right)^2 \xrightarrow{P} 0, \quad s = 2T+1, \dots, 2U. \quad (\text{A.28})$$

By (3.11), (3.14) and (A.17), $\zeta_{2u-1}^\ell = [1 + o_p(1)] D^\ell$ uniformly in ℓ , for $G_u(\mathbf{x}_u) = f_u(\mathbf{x}_u) - f_{u-1}(\mathbf{x}_{u-1})$. Therefore by (A.20),

$$\sum_{\ell=1}^K E_K[(\zeta_{2u-1}^\ell)^2 | \mathcal{F}_{2u-2}] \xrightarrow{P} 0. \quad (\text{A.29})$$

Since $A_{2,T}^\ell = \ell$, $1 \leq \ell \leq K$, (A.29) for $u = T+1$ implies (A.28) for $s = 2T+1$. Since $\zeta_{2u-1}^1, \dots, \zeta_{2u-1}^K$ are independent with mean 0 conditioned on \mathcal{F}_{2u-2} , and ζ_v^ℓ are measurable with respect to \mathcal{F}_{2u-2} for $v \leq 2u-2$,

$$\begin{aligned} & K^{-1} \sum_{j=1}^K E_K \left[\left(\sum_{v=2T+1}^{2u-1} \sum_{\ell: A_{2,t}^\ell = j} \zeta_v^\ell \right)^2 \middle| \mathcal{F}_{2u-2} \right] \\ &= K^{-1} \sum_{j=1}^K \left(\sum_{v=2T+1}^{2u-2} \sum_{\ell: A_{2,t}^\ell = j} \zeta_v^\ell \right)^2 + K^{-1} \sum_{\ell=1}^K E_K[(\zeta_{2u-1}^\ell)^2 | \mathcal{F}_{2u-2}]. \end{aligned}$$

Therefore by (A.29), (A.28) for $s = 2u-1$ follows from (A.28) for $s = 2u-2$.

Since $\text{Var}_K(\#_u^\ell | \mathcal{F}_{2u-1}) = KW_u^\ell(1 - W_u^\ell)$, $\text{Cov}_K(\#_u^i, \#_u^\ell | \mathcal{F}_{2u-1}) = -KW_u^i W_u^\ell$, and χ^ℓ are measurable with respect to \mathcal{F}_{2u-1} ,

$$\begin{aligned} & K^{-1} \sum_{j=1}^K \left[\left(\sum_{v=2T+1}^{2u} \sum_{\ell: A_{2,T}^\ell = j} \zeta_v^\ell \right)^2 \middle| \mathcal{F}_{2u-1} \right] \\ &= K^{-1} \sum_{j=1}^K \left(\sum_{v=2T+1}^{2u-1} \sum_{\ell: A_{2,T}^\ell = j} \zeta_v^\ell \right)^2 + \sum_{\ell=1}^K W_u^\ell (\chi^\ell)^2 - K \left(\sum_{\ell=1}^K W_u^\ell \chi^\ell \right)^2. \end{aligned} \quad (\text{A.30})$$

It follows from an induction argument similar to that used to show (A.29) that

$$\sum_{\ell=1}^k W_u^\ell (\chi^\ell)^2 \xrightarrow{P} 0.$$

Therefore by (A.30), (A.28) for $s = 2u$ follows from (A.28) for $s = 2u - 1$. The induction arguments are complete and we have shown that (A.28) holds for $2T + 1 \leq s \leq 2U$, and Theorem 3 holds. \square

References

- [1] ANDERSON, T. (1984). *An Introduction to Multivariate Statistical Analysis* (2nd edn). Wiley, New York.
- [2] ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods (with discussion). *J. Roy. Statist. Soc. Ser. B* **72** 269–342.
- [3] BILLINGSLEY, P. (1991). *Probability and Measures* (2nd edn). Wiley, New York.
- [4] BRIERS M., DOUCET, A., and SINGH S. S. (2005). Sequential auxiliary particle belief propagation. *8th Intl. Conf. on Inf. Fusion*.
- [5] CAPPÉ, O., MOULINES, É., and RYDEN, T. (2005). *Inference in hidden Markov Models*. Springer: New York.
- [6] CHAN, H.P. and LAI, T.L. (2013). A general theory of particle filters in hidden Markov models with some applications. *Ann. Statist.* **41** 2877–2904.
- [7] CHAN, H.P. and LAI, T.L. (2014). A new approach to Markov chain Monte Carlo with applications to adaptive particle filters. Manuscript.
- [8] CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its applications to Bayesian inference. *Ann. Statist.* **32** 2385–2411.
- [9] CHOPIN, N., JACOB, P.E. and PAPASPILIOPOULOS, O. (2013). SMC²: an efficient algorithm for sequential analysis of state-space models. *J. R. Statist. Soc. Ser. B*, **75**, 397–426.
- [10] DEL MORAL, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, New York.

- [11] DOUC, R. and MOULINES, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36** 2344–2376.
- [12] DOUCET, A. and JOHANSEN, A. M. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. In *Handbook of Nonlinear Filtering* (eds. D. Crisan & B. Rozovsky), Oxford University Press: Oxford.
- [13] FEARNHEAD, P., WYNOLL, D. and TAWN, J. (2010). A sequential smoothing algorithm with linear computational cost. *Biometrika*, **97**, 447–464.
- [14] GORDON, N.J., SALMOND, D.J. and SMITH, A.F.M. (1993). A novel approach to non-linear and non-Gaussian Bayesian state estimation. *IEEE Proc. Radar Signal Processing* **140** 107–113.
- [15] KÜNSCH, H. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021.
- [16] LEE, A., and WHITELEY, N. (2014). Forest resampling for distributed sequential Monte Carlo. arXiv preprint.
- [17] LEE, A., YAU, C., GILES, M., DOUCET, A. and HOLMES, C.C. (2010) On the utility of graphics cards to perform massively parallel implementation of advanced Monte Carlo methods. *J. Comp. Graph. Statist.*, **19**, 769789.
- [18] LINDSTEN, F., JOHANSEN, A. M., NÆSSETH, C. A., KIRKPATRICK, B., SCHON, T. B., ASTON, J. and BOUCHARD, A. (2014). Divide-and-conquer with Sequential Monte Carlo. arXiv preprint.
- [19] PERSING, A. and JASRA, A. (2013). Likelihood estimation for hidden Markov models via generalized two-filter smoothing. *Statist. Prob. Lett.*, **83**, 1433–1442.
- [20] VERGÉ, C., DUBARY, C., DEL MORAL, P. and MOULINES, É. (2014). On parallel implementation of sequential Monte Carlo methods: the island particle model. *Statist. Comp.*, (to appear).

- [21] WHITELEY, N., LEE, A. and HEINE, K. (2013). On the role of interaction in sequential Monte Carlo algorithms. arXiv preprint.